THE JOURNAL OF PHYSICAL CHEMISTRY A

© Copyright 2001 by the American Chemical Society VOLUME 105, NUMBER 2, JANUARY 18, 2001

FEATURE ARTICLE

The Effective Fragment Potential Method: A QM-Based MM Approach to Modeling Environmental Effects in Chemistry

Mark S. Gordon,* Mark A. Freitag, and Pradipta Bandyopadhyay

Department of Chemistry, Iowa State University, Ames, Iowa 50011

Jan H. Jensen and Visvaldas Kairys[‡]

Department of Chemistry, University of Iowa, Iowa City, Iowa 52242

Walter J. Stevens[§]

Physical and Chemical Properties Division, National Institute for Standards and Technology, Gaithersburg, Maryland 20899

Received: July 31, 2000; In Final Form: October 2, 2000

The effective fragment potential (EFP) method is described and its capabilities illustrated using several applications. The original method, EFP1, was primarily developed to describe aqueous solvation, by representing Coulombic, induction and repulsive interactions via one-electron terms in the *ab initio* Hamiltonian. It is demonstrated, using water clusters, the Menshutkin reaction and the glycine neutral/zwitterion equilibrium, that agreement with both fully *ab initio* calculations and experiment are excellent. More recently, the model has been extended so that it can treat any solvent, as well as more difficult links across covalent bonds.

I. Introduction

The primary focus of quantum chemistry has traditionally been on the development and applications of methods that provide increasingly sophisticated descriptions of the behavior and properties of individual molecular species. During the past decade, however, there has been increasing activity in the development of methods whose aim is to understand the effects of the environment (e.g., solvation, heterogeneous catalysis, enzyme activity) on chemical phenomena and more generally in the development of methods for treating condensed phase phenomena. Two alternative types of methods have emerged that are most commonly used to describe environmental effects in chemistry. One type may be collectively referred to as continuum methods.[1] These are characterized by a description of the environment in terms of a single "bulk" unit, in which the identities of individual parts of the environment (e.g., solvent molecules) are not explicitly accounted for. These methods have the advantage that they are *designed* to reproduce bulk or macroscopic behavior, but their description of sometimes important electronic effects is not generally adequate. The second type of method is the discrete approach,[2] in which each individual component of the environment is treated explicitly. To the extent that the potentials that describe these individual

[‡] Current address: Center for Advanced Research in Biotechnology, 9600 Gudelsky Dr., Rockville, MD 20850.

[§] Current address: Chemical Sciences, Geosciences and Biosciences Division, SC-141, Office of Basic Energy Sciences, U.S. Department of Energy, Germantown, MD 20874-1290.

10.1021/jp002747h CCC: $20.00 © 2001 American Chemical Society
Published on Web 12/15/2000

components are reliable, the discrete methods do treat the individual electronic effects, and they attempt to describe bulk behavior by employing techniques such as molecular dynamics or Monte Carlo simulations, combined with the imposition of periodic boundary conditions. The difficulties associated with this second approach are the lack of potentials that are sufficiently sophisticated that they can reproduce both cluster and bulk behavior, and the challenge of adequately sampling configuration space as the number of individual components of the environment (e.g., solvent molecules) increases. It is important to mention that there have been recent efforts designed to incorporate the features of both continuum and discrete methods.[3]

In the past several years, we have been developing a discrete approach that is designed to describe environmental effects. Although this "effective fragment potential" (EFP) method was originally designed[4,5] for the treatment of solvent effects on chemical reactions, it has more recently been used to study clusters of solvent molecules[6,7] and environmental effects in biomolecular systems.[8,9] The most recent developments have been the interface of the method with continuum methods[10] and the extension of the method to the treatment of covalent bonds.[11] In the present paper, we present an overview of the current implementations of the EFP method, in section II. This is followed in section III by a presentation of selected applications, chosen to illustrate the breadth of applicability of the EFP approach. A summary and prognosis for the future are presented in section IV.

II. The Effective Fragment Potential Method

A. Implementation for Water: EFP1. The EFP method was originally formulated[4] and implemented[5] primarily for the purpose of describing clusters of water molecules, either with or without an explicit solute system treated with some level of *ab initio* quantum mechanics. As described in more detail in the following paragraphs, the original method, referred to as EFP1, contains terms that represent (a) Coulombic interactions between solvent molecules and each other, or with the *ab initio* solute, (b) solvent−solvent and solvent−solute induction or polarization interactions and (c) exchange and other repulsion terms. The properties, such as multipoles and polarizabilities, needed to calculate terms (a) and (b) may be determined entirely from *ab initio calculations on a single solvent molecule;* therefore any parameters in these terms may either be established once and stored for future use or generated on the fly as needed. The repulsive term (c), however, arises from explicit intermolecular interactions and thus must either be generated from first principles or obtained by some fitting procedure that is determined by the explicit nature and relative orientation of the intermolecular species (e.g., water dimer) being described. In EFP1, this repulsive term is determined by the latter approach. It is therefore limited either to water, for which the repulsive parameters were obtained, or to cases for which this repulsive term may be omitted.[12]

A schematic of the EFP1 method is presented in Figure 1. The method starts from the *ab initio* Hamiltonian for the solute, which might be a simple molecule, or a reacting group of molecules, plus some (presumably small) number of solvent molecules. The remaining solvent molecules are then treated by adding their effects as one electron terms in the *ab initio* Hamiltonian. At present, the model does not include dynamic correlation. So, these one-electron terms represent the three types of interactions summarized above: Coulombic, induction/polarization, and exchange repulsion + charge transfer. The

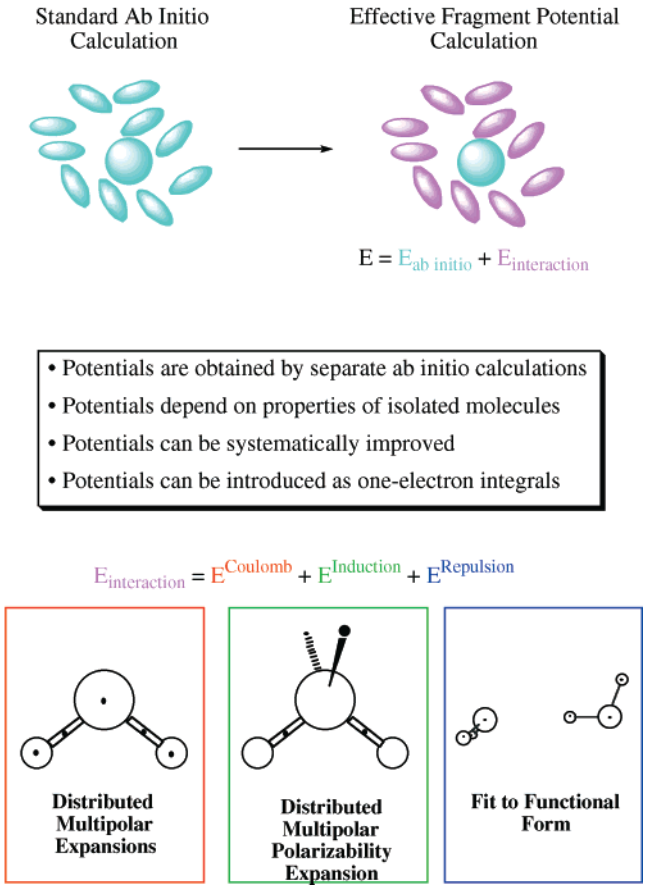


Figure 1. Schematic of the EFP Method.

charge-transfer term is not explicit, but rather is incorporated via the fitting procedure described below. Subsequent extensions of the model, discussed in section IV, will include correlation, and therefore higher order solvent effects (e.g., dispersion, exchange dispersion).

If $\mathbf{H}_{AR}$ is defined as the *ab initio* Hamiltonian that describes the "*ab initio* region" (AR) of the system, then the Hamiltonian $\mathbf{H}$ for the entire system may be written as

$$\mathbf{H} = \mathbf{H}_{AR} + \mathrm{V} \tag{1}$$

The three one-electron terms in $\mathbf{V}$, representing the potential due to the fragment molecules, correspond to electrostatic, polarization, and exchange repulsion/charge-transfer interactions. For the μth solvent molecule, the effective fragment interaction Hamiltonian with an electron in the AR is given by[4,5]

$$V_{el}(\mu,s) = \sum_{k=1}^{K} V_k^{Elec}(\mu,s) + \sum_{l=1}^{L} V_l^{Pol}(\mu,s) + \sum_{m=1}^{M} V_m^{Rep}(\mu,s) \tag{2}$$

where s represents the electronic coordinates. The three terms on the right-hand side (RHS) of eq 2 represent the electrostatic, polarization, and exchange potential/charge-transfer interactions, respectively, and are explained in more detail below. Similar terms are added to represent the interactions between nuclei in the AR and fragment molecules, as well as fragment−fragment interactions. Of course, there are no exchange repulsion/charge transfer terms in the nuclear-fragment interaction. The solute (including the desired number of solvent molecules) is explicitly treated with the *ab initio* wave function of choice, while the remainder may be represented by effective fragments.

An accurate, relatively compact, representation of the electrostatic potential in the important interaction regions is achieved using a *distributed* multipolar analysis[13] (DMA) of the fragment charge distributions. K in the first term of eq 2 is the number of expansion points. In the present implementation, each nuclear center and each bond midpoint is chosen to be an expansion point (e.g., five expansion points for a water molecule), and the expansion is extended through octopoles,

$$V_k^{\mathrm{Elec}}(\mu,s) = \frac{q_k(\mu)q_s}{r_{sk}} - \sum_a^{x,y,z}\mu_a^k(\mu)\,F_a(\mathbf{r}_{sk}) - \frac{1}{3}\sum_{a,b}^{x,y,z}\Theta_{ab}^k(\mu)\,F_{ab}(\mathbf{r}_{sk}) - \frac{1}{15}\sum_{a,b,c}^{x,y,z}\Omega_{abc}^k(\mu)\,F_{abc}(\mathbf{r}_{sk}) \quad (3)$$

where q, μ, Θ and Ω are the charge, dipole, quadrupole and octopole, respectively, and F_a, F_{ab} and F_{abc} are the solute electric field, field gradient and field Hessian. Since the expression in eq 3 is a point charge model, it must be modified to account for overlapping electron densities as the molecules (fragment and solute or fragment and fragment) approach each other. This is accomplished by multiplying the expression in eq 3 by a distance-dependent cutoff function,

$$V_k^{\mathrm{Elec}}(\mu,s) \rightarrow (1-\beta_k(\mu)\mathrm{e}^{-\alpha_k(\mu)r_{sk}^2})V_k^{\mathrm{Elec}}(\mu,s) \quad (4)$$

The polarization of the fragment molecules by the electric field of the *ab initio* molecules (second term in eq 2) is treated by a self-consistent perturbation model employing localized molecular orbitals (LMOs). Using these LMOs, bond and lone pair localized orbital dipole polarizabilities α_{ab}^1 are extracted from finite-field perturbed Hartree−Fock calculations on isolated molecules. For example, if the solvent of interest is water, one would express the total dipole polarizability of water in terms of polarizability tensors obtained for the two OH bond orbitals, two lone pair orbitals and the inner shell orbital. These localized orbital polarizaibilities are centered at the centroids of the L localized molecular orbitals ($L = 5$ for water):

$$V_l^{\mathrm{Pol}}(\mu) = -\sum_{a,b}^{x,y,z}F_a(\mathbf{r}_l)\,\alpha_{ab}^l(\mu)\,\langle F_b(\mathbf{r}_l)\rangle \quad (5a)$$

$$\alpha_{xy}^l = \lim_{F_y\rightarrow 0}\frac{\mu_x^l(F_y)-\mu_x^l(0)}{F_y} \quad (5b)$$

Here, F is the field due to the *ab initio* part of the system, while α_{xy}^l is a component of the polarizability of the fragment molecule in the lth localized orbital.[4]

The exchange-repulsion/charge-transfer interaction between the *ab initio* and fragment molecules is modeled by one-electron terms in the *ab initio* Hamiltonian that have the form of simple Gaussian functions located at the fragment atom centers and the center of mass:

$$V_m^{\mathrm{Rep}}(\mu,s) = \sum_j^J \beta_{m,j}(\mu)\mathrm{e}^{-\alpha_{m,j}(\mu)r_{m,s}^2} \quad (6)$$

For water, $J = 2$, and there are four fragment centers ($M = 4$)—the three nuclear centers and the center of mass. The Gaussian functions are optimized by a fitting procedure. *Ab initio* calculations are performed on some number of points (e.g., 192 for water dimer to represent water as the solvent). The *ab initio* exchange repulsion/charge-transfer potential is obtained by subtracting the sum of electrostatic plus polarization energies from the total potential to obtain the term $E_{\mathrm{rem}}^{(ab)}$. Then, V_m^{Rep} is fitted to $E_{\mathrm{rem}}^{(ab)}$:

$$\Delta = \sum_p^P w_p[\langle\Psi|\sum_m^M V_m^{\mathrm{Rep}}|\Psi\rangle_p - E_{\mathrm{rem}}^{(ab)}(p)]^2 \quad (7)$$

where w_p is a weighting factor that has been set to unity in the fitting process for water. Ψ is the wave function of the AR. The fragment−fragment interactions are treated in a similar manner, except that a single exponential, rather than Gaussian functions, is used to represent the exchange-repulsion/charge-transfer interaction.

The necessary equations have been derived[4,5] and coded for the analytic gradients of the entire (*ab initio* + fragments) system. The availability of analytic gradients means that one can also perform vibrational analyses using finite differences of these gradients. So, one can determine the manner in which geometries and energetics of minima, transition states, and reaction paths (and therefore the reaction dynamics) are modified by the presence of the solvent. The entire code described here has been added to the electronic structure code GAMESS.[14]

B. General Implementation: EFP2. The method described in section II.A can in principle be extended to any solvent. However, such a generalization would require extensive calculations on the dimer for any solvent of interest, at many distances and orientations, to obtain a repulsive potential that can be fitted to a functional form as in eqs 6 and 7. This process would be very time-consuming and not necessarily successful. A desirable alternative approach would be to replace the fitted exchange repulsion/charge-transfer term with expressions that are derived from first principles. This more general method may still be described by the schematic in Figure 1, with the fitted repulsive potential replaced by separate terms for exchange repulsion and charge transfer. It is also necessary to ensure that the generalized fragment−fragment interaction accounts properly for the penetration of overlapping charge densities, represented by the screening term in eq 4 for EFP1. This general method, referred to here as EFP2, is a work in progress[15−17] that is summarized in this section.

1. Exchange Repulsion. The exact zeroth order exchange repulsion energy between wave functions Ψ_A and Ψ_B (assumed here to be RHF wave functions) may be extracted from the Heitler−London energy by subtracting the classical Coulomb energy as well as the energies of molecules A and B,

$$E_{\mathrm{exch}} = \frac{\langle\Psi_A\Psi_B|\mathbf{A}H_{AB}|\Psi_A\Psi_B\rangle}{\langle\Psi_A\Psi_B|\mathbf{A}\Psi_A\Psi_B\rangle} - \langle\Psi_A\Psi_A|V_{AB}|\Psi_B\Psi_B\rangle - E_A - E_B \quad (8)$$

Here H_{AB} is the super molecular Hamiltonian for the A−B complex, given by the individual Hamiltonians for molecules A and B plus the interaction operator,

$$H_{AB} = H_A + H_B + V_{AB} \quad (9)$$

Using this division for H_{AB} we can express the exchange repulsion energy in terms of two internal energy contributions as well as an interaction term,

$$E_{\text{exch}} = \frac{\langle \Psi_A \Psi_B | \mathbf{A}(H_A + H_B) | \Psi_A \Psi_B \rangle}{\langle \Psi_A \Psi_B | \mathbf{A} \Psi_A \Psi_B \rangle} - E_A - E_B + \frac{\langle \Psi_A \Psi_B | \mathbf{A} V_{AB} | \Psi_A \Psi_B \rangle}{\langle \Psi_A \Psi_B | \mathbf{A} \Psi_A \Psi_B \rangle} - \langle \Psi_A \Psi_A | V_{AB} | \Psi_B \Psi_B \rangle$$

$$= \Delta E_A + \Delta E_B + E_{\text{exch}}(V) \quad (10)$$

The exchange repulsion energy arises from terms in the wave function generated by the antisymmetrizer, **A**, which permutes 0, 1, 2, ... electron pairs,

$$\mathbf{A} = 1 - P_1 + P_2 - ... \quad (11)$$

In the EFP2 method, the expansion in eq 11 is truncated at P_1. This leads to an approximate exchange repulsion that is proportional to the square of the *intermolecular* overlap (S),

$$\mathbf{A} \approx 1 - P_1 \Rightarrow E_{\text{exch}} \approx E_{\text{exch}}[O(S^2)] \quad (12)$$

where (cf. eq 10),

$$E_{\text{exch}}[O(S^2)] = \Delta E_A[O(S^2)] + \Delta E_B[O(S^2)] + E_{\text{exch}}[V;O(S^2)] \quad (13)$$

For exact HF wave functions the internal energy contribution to the second-order exchange energy vanishes,

$$\Delta E_A[O(S^2)] \approx 0 \quad (14)$$

and similarly for $\Delta E_B[O(S^2)]$. In practice, the HF MOs are expanded in a finite basis set, so neglecting the internal energy contributions in eq 13,

$$E_{\text{exch}}[O(S^2)] \approx E_{\text{exch}}[V;O(S^2)] \quad (15)$$

introduces a basis set dependent approximation. This expression for the exchange repulsion energy can be separated into three distinct energy terms based on their *explicit* overlap dependence.

$$E_{\text{exch}}[V;O(S^2)] = E_{\text{exch}}(S^0) + E_{\text{exch}}(S^1) + E_{\text{exch}}(S^2) \quad (16)$$

That is, all three terms *scale* as S^2, but their explicit dependence on the overlap is zeroth, first and second order, respectively. Each term is then approximated separately in terms of localized molecular orbitals (LMOs):

$$E_{\text{exch}}(S^0) = -2\sum_{i\in A}\sum_{j\in B}\langle i|K_j|i\rangle \approx -2\sum_{i\in A}\sum_{j\in B} 2\sqrt{\frac{-2\ln S_{ij}}{\pi}}\frac{S_{ij}^{\ 2}}{R_{ij}} \quad (17)$$

$$E_{\text{exch}}(S^1) = -2\sum_{i\in A}\sum_{j\in B} S_{ij}[V_{ij,A} + V_{ij,B} + \sum_{k\in A}\langle i|2J_k - K_k|j\rangle + \sum_{l\in B}\langle i|2J_l - K_l|j\rangle]$$

$$\{= -2\sum_{i\in A}\sum_{j\in B} S_{ij}[F_{ij}^A + F_{ij}^B - 2T_{ij}]\}$$

$$\approx -2\sum_{i\in A}\sum_{j\in B} S_{ij}[\sum_{k\in A} F_{ik}^A S_{kj} + \sum_{l\in A} F_{jl}^B S_{li} - 2T_{ij}] \quad (18)$$

$$E_{\text{exch}}(S^2) = 2\sum_{i\in A}\sum_{j\in B} S_{ij}[\sum_{k\in A} S_{kj}(V_{ik,B} + \sum_{l\in B}\langle i|2J_l|k\rangle) + \sum_{l\in B} S_{il}(V_{jl,A} + \sum_{k\in A}\langle j|2J_k|l\rangle) - \sum_{k\in A}\sum_{l\in B} S_{kl}\langle ik|lj\rangle]$$

$$\approx \{2\sum_{i\in A}\sum_{j\in B} S_{ij}^{\ 2}[V_{ii,B} + \sum_{l\in B}\langle i|2J_l|i\rangle + V_{jj,A} + \sum_{k\in A}\langle j|2J_k|j\rangle - \langle i|J_j|i\rangle]\}$$

$$\approx 2\sum_{i\in A}\sum_{j\in B} S_{ij}^{\ 2}[\sum_{J\in B} -Z_J R_{iJ}^{\ -1} + 2\sum_{l\in B} R_{il}^{\ -1} + \sum_{I\in A} -Z_I R_{Ij}^{\ -1} + 2\sum_{k\in A} R_{kj}^{\ -1} - R_{ij}^{\ -1}] \quad (19)$$

S_{ij} and T_{ij} are, respectively, an overlap and kinetic energy integral connecting LMOs i and j. R_{ij} is the distance between the centroids of charge of LMOs i and j, and R_{ij} is the distance between nucleus I (with nuclear charge Z_I) and the centroid of charge of LMO j. F_{ik}^A is the Fock matrix element connecting LMO i with LMO k within molecule A; i.e., it is an intrafragment Fock matrix element. The approximations in eq 12 and eq 14 are based on the assumption that LMOs are used. Combining these three approximations for the zeroth, first and second-order terms in the intermolecular overlap leads to the approximate formula for exchange repulsion between closed shell molecules.[15]

$$E_{\text{exch}} \approx -2\sum_{i\in A}\sum_{j\in B} 2\sqrt{\frac{-2\ln S_{ij}}{\pi}}\frac{S_{ij}^{\ 2}}{R_{ij}} - 2\sum_{i\in A}\sum_{j\in B} S_{ij}[\sum_{k\in A} F_{ik}^A S_{kj} + \sum_{l\in B} F_{jl}^B S_{li} - 2T_{ij}] + 2\sum_{i\in A}\sum_{j\in B} S_{ij}^{\ 2}[\sum_{J\in B} -Z_j R_{iJ}^{\ -1} + 2\sum_{l\in B} R_{il}^{\ -1} + \sum_{I\in A} -Z_I R_{Ij}^{\ -1} + 2\sum_{k\in A} R_{kj}^{\ -1} - R_{ij}^{\ -1}] \quad (20)$$

Note that the intramolecular Fock matrix elements in eq 20 only need to be calculated once for a particular molecular species and a particular atomic basis set.

The accuracy and CPU cost of the exchange energies calculated according to eq 20 are summarized in Table 1, for several dimers calculated using three different basis sets. The errors relative to the exact exchange energies are reasonably small for all basis sets, but clearly improve (decrease) for the larger basis sets. The CPU requirements for the EFP2 method are a very small fraction of the corresponding *ab initio* CPU times.

Note that eq 20 applies rigorously only to cases in which all MOs correspond to those of free monomers A and B and are localized. This complicates the application of eq 20 to *ab initio*/EFP interactions, in which one of the molecules (A) is treated by an *ab initio* wave function, since the MOs are optimized in canonical form and in the presence of B. The complication arises due to the orthogonality "tails" that occur in the *ab initio* treatment, even for localized orbitals. One solution to this complication is the following: First, the *ab initio* MOs on A are optimized using the Hamiltonian of eq 1, including the first two terms of eq 2 to represent the field of B. These *ab initio* MOs are localized and then introduced into eq 20. While this approach gives reasonable exchange repulsion energies, charge-transfer effects are ignored, and the exact gradients require the solution of a coupled Hartree−Fock equation, as discussed below. A solution to both problems is to derive and implement an exchange repulsion *Fock operator* by taking the variational derivative of eq 16 and replacing the resulting two-electron

TABLE 1: Exact (EFP) Exchange Repulsion Energies and Requisite CPU Times for an Energy and Gradient Evaluation Calculated for Several Dimers Using Three Basis Sets[a]

	basis set		
	6-31+G(d,p)	6-31++G(2d,2p)	pVTZ
	Water Dimer		
E_{exch}	4.76 (4.41)	4.78 (4.90)	4.70 (4.88)
CPU seconds	52 (0.4)	251 (0.5)	839 (0.7)
	Methanol Dimer		
E_{exch}	5.07 (4.43)	5.19 (5.23)	5.22 (5.63)
CPU seconds	792 (0.8)	3476 (1.0)	12851 (1.6)
	Dichloromethane Dimer		
E_{exch}	0.80 (0.30)	0.84 (0.35)	0.84 (0.60)
CPU seconds	2226 (1.0)	14014[b] (1.1)	27222[b] (1.5)
	Acetonitrile Dimer		
E_{exch}	2.04 (1.60)	2.12 (1.94)	2.02 (1.96)
CPU seconds	1754 (1.0)	14608[b] (1.3)	26499[b] (2.1)
	Acetone Dimer		
E_{exch}	2.12 (1.53)	2.19 (1.72)	2.05 (1.67)
CPU seconds	6376 (1.6)	58519[b] (2.1)	602781[b] (3.9)
	DMSO Dimer		
E_{exch}	7.27 (6.59)	7.58 (7.42)	7.67 (8.65)
CPU seconds	20807[b] (2.1)	115155[b] (2.4)	671950[b] (4.5)

[a] Energies are in kcal/mol. [b] SCF calculations run in direct mode due to size.

integrals with approximations similar to those outlined in eqs 17−19. The derivation of this operator as well as the EFP/EFP exchange induction and charge-transfer energies will be presented in a forthcoming paper.

2. Charge Penetration. The electrostatic interaction between two molecules separated by a large distance is well represented by the expressions for classical multipolar interactions. However, if the two molecules are brought close enough, such that their charge densities overlap, the nuclei on one molecule will no longer be shielded by its own electron density, and will experience a greater attraction for the electron density associated with the other species. The energy difference resulting from this increased attraction is the charge penetration.

Consider the interaction between a hydrogen-like atom with nuclear charge Z and a proton. The wave function of the former is given by[13]

$$\psi(r) = \left(\frac{Z^3}{\pi}\right)^{1/2} e^{-Zr} \tag{21}$$

with corresponding electron charge density

$$\rho(r) = -\frac{Z^3}{\pi} e^{-2Zr} \tag{22}$$

One can then use Poisson's equation ($\nabla^2 V = -\rho/\epsilon_o$), where ϵ_o is the permittivity of free space, to find the potential due to that density. This results in

$$V(r) = -\frac{1}{r} + e^{-2Zr}\left(Z + \frac{1}{r}\right) \tag{23}$$

The second term in eq 23, the charge penetration, falls off as a simple exponential.

One way to calculate charge penetration between fragments is to introduce a damping function that multiplies the electrostatic potential. Equation 23 can be rewritten as

$$V(r) = [1 - e^{-2Zr}(1 + rZ)]\left[-\frac{1}{r}\right] = f^{\,\mathrm{damp}}(r)\, V^{\mathrm{mult}}(r) \tag{24}$$

This suggests that a multipole expansion of the electrostatic potential (V^{mult}) can be corrected for charge penetration effects by using a damping function, $f^{\,\mathrm{damp}}$.

It is important choose the parameter in the damping function such that the function fits the molecular *ab initio* electrostatic potential well in the region of interest. Then the difference between the damped and undamped electrostatic interactions, within the framework of the distributed multipolar analysis (DMA), will be a good approximation to the charge penetration.

Consider two charge densities ρ_A and ρ_B centered at points A and B, respectively. These points represent atomic centers and bond midpoints for EFPs. Points 1 and 2 represent electronic positions associated with ρ_A and ρ_B, respectively, referenced from an arbitrary origin, O. Using these definitions, the electrostatic interaction of the two charge densities is given by

$$\begin{aligned} E^{\mathrm{Elec}} &= \int\int d\mathbf{r}_1\, d\mathbf{r}_2\, \rho_A(\mathbf{r}_{1A})\, \rho_B(\mathbf{r}_{2B}) |\mathbf{r}_{12}|^{-1} \\ &= \int d\mathbf{r}_1\, \rho_A(\mathbf{r}_{1A}) \int d\mathbf{r}_2\, \rho_B(\mathbf{r}_{2B}) |\mathbf{r}_{12}|^{-1} \end{aligned} \tag{25}$$

where $\mathbf{r}_{1A} = \mathbf{r}_1 - \mathbf{R}_A$. In the EFP method the electrostatic potential due to the charge density is expanded in terms of charges, dipoles, quadrupoles, and octopoles at each atomic center and bond midpoint using Stone's distributed multipolar analysis:

$$\begin{aligned} E^{\mathrm{Elec}} &= \int d\mathbf{r}_1\, \rho_A(\mathbf{r}_{1A}) \int d\mathbf{r}_2\, \rho_B(\mathbf{r}_{2B}) \left[|\mathbf{r}_{1B}|^{-1} - \frac{(\mathbf{r}_{12} - \mathbf{r}_{1B})\mathbf{r}_{1B}}{r_{1B}^{\,3}} + \ldots \right] \\ &= \int d\mathbf{r}_1\, \rho_A(\mathbf{r}_{1A}) V_B^{\mathrm{mult}}(\mathbf{r}_{1B}) \end{aligned} \tag{26}$$

where V_B^{mult} is expanded in multipolar terms:

$$V_B^{\mathrm{mult}}(\mathbf{r}_{1B}) = V_B^{\mathrm{charge}}(\mathbf{r}_{1B}) + V_B^{\mathrm{dipole}}(\mathbf{r}_{1B}) + \ldots \tag{27}$$

The effect of charge penetration is accounted for by multiplying eq 27 by a damping function that satisfies the following requirements: (a) go to unity for large R_{AB} and fall off toward zero as R_{AB} (the distance between points A and B) approaches zero, (b) fit well to the *ab initio* electrostatic potential of an isolated fragment in the region of its Van der Waals radius, and (c) give rise to tractable integrals in eq 26. A simple exponential function provides the best balance of the desired qualities:[18]

$$\tilde{V}_A^{\mathrm{mult}}(\mathbf{r}_{1A}) = (1 - e^{-\alpha_A r_{1A}}) V_A^{\mathrm{mult}}(\mathbf{r}_{1A}) \tag{28}$$

The parameter α is determined by minimizing the difference, Δ, between the quantum mechanical electrostatic potential (ES) and the multipolar expansion of the potential over a grid of points (see eq 39 below).

To account for the fact that two damped distributed multipolar expansions are interacting, the charge density on A, $\rho_A(\mathbf{r}_{1A})$, is found by applying Poisson's equation to the damped charge potential, eq 28:

$$\rho_A(\mathbf{r}_{1A}) = -\epsilon_o \nabla^2 \tilde{V}_A^{\mathrm{mult}}(\mathbf{r}_{1A}) = \rho_A^{\mathrm{charge}}(\mathbf{r}_{1A}) + \rho_A^{\mathrm{dipole}}(\mathbf{r}_{1A}) + \ldots \tag{29}$$

Since Poisson's equation is applied to each term in the damped charge electrostatic potential, the charge density is also expressed in terms of charge, dipole, quadrupole, etc. contributions. Then using eqs 27 and 29, the integral in eq 26 becomes

$$\begin{aligned}\tilde{E}^{\mathrm{Elec}} &= \int \mathrm{d}\mathbf{r}_1\, \rho_{\mathrm{A}}(\mathbf{r}_{1\mathrm{A}})\, \tilde{V}_{\mathrm{B}}^{\mathrm{mult}}(\mathbf{r}_{1\mathrm{B}}) \\ &= \int \mathrm{d}\mathbf{r}_1\, [\rho_{\mathrm{A}}^{\mathrm{charge}}(\mathbf{r}_{1\mathrm{A}}) + \rho_{\mathrm{A}}^{\mathrm{dipole}}(\mathbf{r}_{1\mathrm{A}}) + \\ &\qquad ...][\tilde{V}_{\mathrm{B}}^{\mathrm{charge}}(\mathbf{r}_{1\mathrm{B}}) + \tilde{V}_{\mathrm{B}}^{\mathrm{dipole}}(\mathbf{r}_{1\mathrm{B}}) + ...] \quad (30)\end{aligned}$$

Note that each integral in eq 30 must be explicitly symmetrized, since, for example, the charge density on A interacting with the damped charge potential on B is not the same as the charge density on B interacting with the damped charge potential on A.[18] For simplicity, only one-half of the symmetrized integrals will be explicitly discussed.

From eqs 27−29 one finds

$$\rho_{\mathrm{A}}^{\mathrm{charge}}(\mathbf{r}_{1\mathrm{A}}) = \frac{q_{\mathrm{A}}\alpha_{\mathrm{A}}{}^2\epsilon_{\mathrm{o}}}{r_{1\mathrm{A}}}\mathrm{e}^{-\alpha_{\mathrm{A}}r_{1\mathrm{A}}} \quad (31)$$

$$\tilde{V}_{\mathrm{B}}^{\mathrm{charge}}(\mathbf{r}_{1\mathrm{B}}) = [1 - \mathrm{e}^{-\alpha_{\mathrm{B}}r_{1\mathrm{B}}}]q_{\mathrm{B}}r_{1\mathrm{B}}{}^{-1} \quad (32)$$

where the charge at point A, q_{A}, is found using Stone's method[13] and ϵ_{o} is the permittivity of free space. Then, the first term in eq 30 becomes

$$\tilde{E}_{\mathrm{chg-chg}}^{\mathrm{Elec}} \approx q_{\mathrm{A}}q_{\mathrm{B}}\alpha_{\mathrm{A}}{}^2\epsilon_{\mathrm{o}}\left[\int \mathrm{d}\mathbf{r}_1 \frac{\mathrm{e}^{-\alpha_{\mathrm{A}}r_{1\mathrm{A}}}}{r_{1\mathrm{A}}r_{1\mathrm{B}}} - \int \mathrm{d}\mathbf{r}_1 \frac{\mathrm{e}^{-\alpha_{\mathrm{A}}r_{1\mathrm{A}}}\,\mathrm{e}^{-\alpha_{\mathrm{B}}r_{1\mathrm{B}}}}{r_{1\mathrm{A}}r_{1\mathrm{B}}}\right] \quad (33)$$

which can be evaluated to yield

$$\tilde{E}_{\mathrm{chg-chg}}^{\mathrm{Elec}} \approx \frac{q_{\mathrm{A}}q_{\mathrm{B}}}{R_{\mathrm{AB}}}\left[1 - \mathrm{e}^{-\alpha_{\mathrm{A}}R_{\mathrm{AB}}} - \frac{\alpha_{\mathrm{A}}{}^2}{(\alpha_{\mathrm{A}}{}^2 - \alpha_{\mathrm{B}}{}^2)}[\mathrm{e}^{-\alpha_{\mathrm{B}}R_{\mathrm{AB}}} - \mathrm{e}^{-\alpha_{\mathrm{A}}R_{\mathrm{AB}}}]\right] \quad (34)$$

for $\alpha_{\mathrm{A}} \neq \alpha_{\mathrm{B}}$, and

$$\tilde{E}_{\mathrm{chg-chg}}^{\mathrm{Elec}} \approx \frac{q_{\mathrm{A}}q_{\mathrm{B}}}{R_{\mathrm{AB}}}\left[1 - \mathrm{e}^{-\alpha R_{\mathrm{AB}}}\left(1 + \frac{\alpha R_{\mathrm{AB}}}{2}\right)\right] \quad (35)$$

for $\alpha_{\mathrm{A}} = \alpha_{\mathrm{B}} = \alpha$, since $4\pi\epsilon_{\mathrm{o}} = 1$ in atomic units. A similar procedure is used to calculate electron−nuclear interactions, with the result given in eq 36:

$$\begin{aligned}E_{\mathrm{chg}}^{\mathrm{Elec-Nuc}} &= \int \mathrm{d}\mathbf{r}_1\, \rho_{\mathrm{A}}^{\mathrm{charge}}(\mathbf{r}_{1\mathrm{A}})Z_{\mathrm{B}}r_{1\mathrm{B}}{}^{-1} + \\ &\qquad \int \mathrm{d}\mathbf{r}_1\, \rho_{\mathrm{B}}{}^{\mathrm{charge}}(\mathbf{r}_{1\mathrm{B}})Z_{\mathrm{A}}r_{1\mathrm{A}}{}^{-1} \\ &= q_{\mathrm{A}}Z_{\mathrm{B}}\alpha_{\mathrm{A}}{}^2\int \mathrm{d}\mathbf{r}_1 \frac{\mathrm{e}^{-\alpha_{\mathrm{A}}r_{1\mathrm{A}}}}{r_{1\mathrm{A}}r_{1\mathrm{B}}} + q_{\mathrm{B}}Z_{\mathrm{A}}\alpha_{\mathrm{B}}{}^2\int \mathrm{d}\mathbf{r}_1 \frac{\mathrm{e}^{-\alpha_{\mathrm{B}}r_{1\mathrm{B}}}}{r_{1\mathrm{A}}r_{1\mathrm{B}}} \\ &= \frac{q_{\mathrm{A}}Z_{\mathrm{B}}}{R_{\mathrm{AB}}}[1 - \mathrm{e}^{-\alpha_{\mathrm{A}}R_{\mathrm{AB}}}] + \frac{q_{\mathrm{B}}Z_{\mathrm{A}}}{R_{\mathrm{AB}}}[1 - \mathrm{e}^{-\alpha_{\mathrm{B}}R_{\mathrm{AB}}}] \quad (36)\end{aligned}$$

Summing eqs 35 and 36, including the symmetrization, and subtracting out the undamped interactions, the charge−charge contribution to the charge penetration energy is

$$\begin{aligned}E_{\mathrm{chg-chg}}^{\mathrm{Pen}} = -\frac{1}{2R_{\mathrm{AB}}}\Big[&q_{\mathrm{A}}(q_{\mathrm{B}} + 2Z_{\mathrm{B}})\mathrm{e}^{-\alpha_{\mathrm{A}}R_{\mathrm{AB}}} + \\ &q_{\mathrm{B}}(q_{\mathrm{A}} + 2Z_{\mathrm{A}})\mathrm{e}^{-\alpha_{\mathrm{B}}R_{\mathrm{AB}}} + \frac{q_{\mathrm{A}}q_{\mathrm{B}}(\alpha_{\mathrm{A}}{}^2 + \alpha_{\mathrm{B}}{}^2)}{\alpha_{\mathrm{A}}{}^2 - \alpha_{\mathrm{B}}{}^2}(\mathrm{e}^{-\alpha_{\mathrm{B}}R_{\mathrm{AB}}} - \mathrm{e}^{-\alpha_{\mathrm{A}}R_{\mathrm{AB}}})\Big]\end{aligned} \quad (37)$$

for $\alpha_{\mathrm{A}} \neq \alpha_B$ and $Z_{\mathrm{A,B}} = 0$ for bond midpoints. For $\alpha_{\mathrm{A}} = \alpha_{\mathrm{B}} = \alpha$,

$$E_{\mathrm{chg-chg}}^{\mathrm{Pen}} = -\frac{1}{R_{\mathrm{AB}}}\left[q_{\mathrm{A}}q_{\mathrm{B}}\left(1 + \frac{\alpha R_{\mathrm{AB}}}{2}\right) + q_{\mathrm{A}}Z_{\mathrm{B}} + q_{\mathrm{B}}Z_{\mathrm{A}}\right]\mathrm{e}^{-\alpha R_{\mathrm{AB}}} \quad (38)$$

In eqs 37 and 38 the total charge penetration is the sum of all charge penetration energies between unique pairs of intermolecular DMA points A and B.

If one follows the above procedures for the next higher terms in the expansion, analytically solvable integrals are obtained for the charge−dipole interactions, but not for the dipole-charge terms.[18] Therefore, as a first approximation, only the charge−charge interactions are retained. Fortunately, most of the total charge penetration is still recovered.

The parameter α in the damping function is obtained using the error function

$$\Delta = \sum_{\text{grid points}} [V_{\text{ab initio}}^{\mathrm{ES}} - V_{\text{damped multipole}}^{\mathrm{ES}}]^2 \quad (39)$$

based on the difference between the *ab initio* and multipolar electrostatic potentials. A grid is defined about an isolated fragment molecule by placing concentric spheres about each of the atom centers at 67% and 300% of the van der Waals radius of the corresponding atom. These values were chosen because they result in the best fit of the damping function to the *ab initio* density, and they describe the physically most important regions in terms of charge penetration.[18] The fragment is then placed within a three-dimensional Cartesian grid with a spacing of 0.50 Bohr in each direction, and any point not within the two spheres is discarded. The *ab initio* density is calculated on the fragment during a GAMESS run, and the electrostatic potential is computed at each grid point. The parameter α is optimized in the exponential damping function such that Δ in eq 39 is minimized. The average absolute difference between the EFP method and the *ab initio* charge penetration for dimers of water, methanol, acetonitrile, acetone, DMSO, and dichloromethane at their equilibrium geometry is 0.32 kcal/mol.

An alternative, more CPU-intensive, expression for E^{pen}, may be derived[19] using an approach analogous to that used to derive eq 16, the Spherical Gaussian Overlap (SGO) approximation:

$$E^{\mathrm{pen}} = -2\sum_{i\in\mathrm{A}}\sum_{j\in\mathrm{B}}(-2\ln S_{ij})^{-1/2}\frac{S_{ij}{}^2}{R_{ij}} \quad (40)$$

In eq 40 S_{ij} is the intermolecular overlap integral between molecular orbitals i and j. The accuracy of this approach again depends on the use of localized molecular orbitals.

3. Energy Gradients. Analytic expressions for the energy first derivatives with respect to the nuclear coordinates for the EFP1 method were presented in the original paper.[4] This derivation was extended to minimum energy path (MEP) calculations in a subsequent paper,[5] so one can perform any type of calculation that requires first, or even numerical second, derivatives.

Analytical expressions for the energy first derivatives for EFP2, eq 20, with respect to EFP coordinates were presented in ref 16. As mentioned above, the corresponding derivatives with respect to *ab initio* coordinates are complicated by the fact that the exchange repulsion effect is not added as an operator as in eq 2. As a result, the current formulation requires the use of coupled perturbed Hartree−Fock equations. Since this is very time-consuming, the derivation of an operator representation of eq 16 is the preferred alternative.

C. Extension to Covalent Bonds. The EFP method was originally developed to study the weak interactions between separate molecules, such as solvent−solvent and solvent−solute interactions. Since the fragments are represented by model potentials (EFPs), the method may be considered to be in the general category of QM/MM (quantum mechanics/molecular mechanics) methods. In other contexts QM/MM methods have also been very useful for describing extended systems in which the QM and MM regions are separated by covalent bonds rather than weak intermolecular forces. Applications include the study of organometallic compounds[20] and surface chemistry.[21] Most of these methods rely on somewhat arbitrary links between the QM and MM regions.

To make the link between the *ab initio* and MM portions of a problem less arbitrary, a covalently bonded *ab initio*/EFP interface has recently been developed[11] and implemented in GAMESS.[14] The method is similar in spirit to that of Assfeld and Rivail.[22] The essential features of the approach are as follows:

(1) A buffer region consisting of several LMOs, typically surrounding the α-carbon of a given side-chain, is defined as the *ab initio*/EFP boundary. Once the buffer region is defined, these LMOs are obtained by an *ab initio* calculation on all or a subset of the system (see below for an example), projected onto the buffer atom basis functions.[23] These LMOs are subsequently frozen in the EFP calculations by setting select MO Fock matrix elements to zero.[24,25] The *ab initio*/buffer region interactions are calculated by including the exact quantum mechanical Coulomb and exchange operators corresponding to the charge distribution of the buffer region, in the *ab initio* Hamiltonian. This requires calculation of two-electron integrals over basis functions in the buffer region. Since the buffer MOs are frozen, the changes in induction (polarization) contributions from the buffer region are neglected during a geometry optimization of the *ab initio* region. The effect of this approximation on the chemical reaction of interest can be systematically reduced by increasing the size of the *ab initio* region.

(2) Variational collapse of the *ab initio* wave function into the buffer region is avoided by keeping the *ab initio* MOs orthogonal to the buffer LMOs by Schmidt orthogonalization. This is an approximation relative to a full *ab initio* calculation because the MOs are allowed to build up "orthogonality tails" only in the buffer region, not in the EFP region. The associated error can again be systematically reduced by increasing the size of the buffer region.

(3) The remaining part of the system (or within a defined radius of the active region) is represented by an EFP. The presence of the buffer region provides sufficient separation between the EFP and the *ab initio* regions to ensure that the remaining interactions can be treated as nonbonded interactions via the EFP terms presented above.

We demonstrate the utility of this method by calculating the proton affinity of the ϵ-NH_2 group in the amino acid lysine and the Gly-Lys-Gly tripeptide. These molecules are small enough to make full *ab initio* calculations feasible, but large enough to allow for several different choices of buffer region. Figure 2 depicts the general scheme of our method using lysine, divided into the following EFP/buffer/*ab initio* regions $(CO_2H)(NH_2)CH-/CH_2CH_2/-CH_2CH_2NH_2$, as an example.

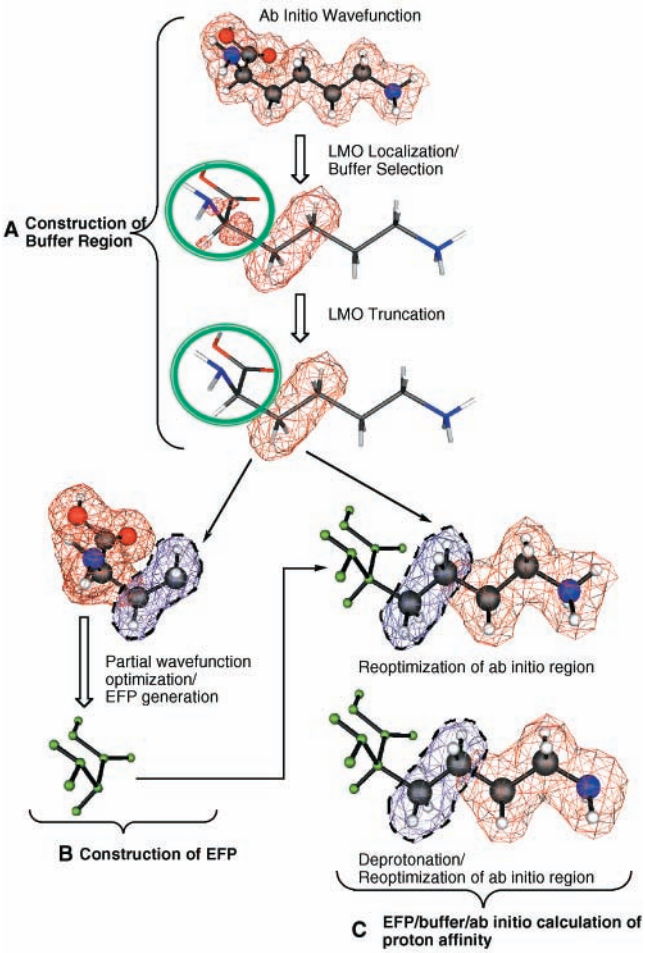


Figure 2. Schematic representation of buffer and EFP generation.

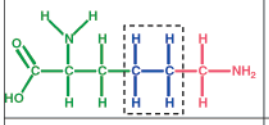
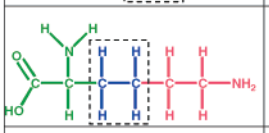
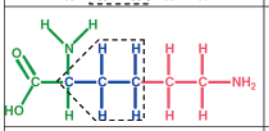
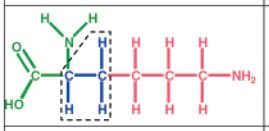
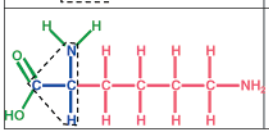
The RHF/6-31G* optimized structure of protonated lysine ($LysH^+$) is obtained and the MOs are localized using the Edmiston-Ruedenberg localization scheme.[26] The LMOs that will comprise the buffer are selected and projected, using the corresponding orbital method, so that they only span basis functions on the atoms in the buffer region. The best source of buffer LMOs is presumably the LMOs calculated for the entire molecule, or the largest possible piece thereof.

The density of the molecular region that will be described by the EFP is re-optimized in the presence of the buffer region but in the absence of the *ab initio* region. The electrostatic potential of the optimized density, but not the buffer density, is expanded in terms of multipoles through octopoles centered at all atomic and bond midpoints using Stone's Distributed Multipole Analysis.[13] Calculated in this way, these multipoles do not account for polarization of the EFP region due to the *ab initio* region, so that this effect is not double counted when dipole polarizabilities are added.

The EFP, buffer, and *ab initio* regions are combined for $LysH^+$ and the geometry of the *ab initio* region is re-optimized. In a second calculation the proton is removed and the *ab initio* region geometry is re-optimized. The energy difference between these two systems is taken to be the proton affinity.

Table 2 lists the PAs calculated with buffer regions (constructed as outlined above) at increasingly larger distances from

TABLE 2: Proton Affinities of Lysine (kcal/mol)[a]

	1 Reference ab initio calculation	2 QM/buffer calculation	3 QM/buffer/EFP calculation
	237.3 0.0	239.7 +2.4	239.5 +2.2
	237.3 0.0	236.9 –0.4	237.8 +0.5
	237.3 0.0	237.1 –0.2	237.2 –0.1
	237.2 0.0	234.6 –2.6	237.0 –0.2
	237.2 0.0	237.2 0.0	Blows up

[a] The upper number is the absolute proton affinity, the lower one is the error relative to the reference *ab initio* calculation in column 1. The proton affinity of the fully relaxed lysine is 236.6 kcal/mol.

the $\epsilon-$N. The buffer regions, in bold and boxed in Table 2, represent LMOs calculated for the RHF/6-31G* optimized structure of protonated lysine, and truncated by projection. The buffer region in the first row of the table, for example, consists of four CH bonds, one CC bond, and two C 1s core MOs. The geometry of the buffer and EFP regions (always to the left of the buffer region) is thus taken from the RHF/6-31G* optimized geometry of protonated lysine.

The target value is the PA calculated using fully relaxed RHF/6-31G* wave functions and geometries of the protonated and unprotonated form for lysine, 236.6 kcal/mol. It is important to separate the error introduced by approximating part of the electronic charge distribution with an EFP and buffer, from the error introduced by the geometric constraints on those regions. This is accomplished by also calculating the PA using RHF/6-31G* for the entire molecule, but only partial geometry optimization with the same geometrical constraints as in the EFP calculations. The values are listed in the first column of Table 2, and show that the geometrical constraints introduce an error of 0.6 kcal/mol. In the subsequent discussion, we take these constrained all-*ab-initio* calculations as our reference for the corresponding EFP calculations.

Column 3 of Table 2 lists the results obtained for the EFP/buffer/*ab initio* calculation. It is evident that the PA converges relatively quickly to within 0.2 kcal/mol of the all-*ab-initio* reference value. Column 2 lists the corresponding PA value without the EFP to isolate the effect of the EFP region of the molecule on the PA, which can be as large as 2.6 kcal/mol for this system. The C_a-C_b bond and the associated CH and core LMOs ([$\alpha\beta$]-buffer) appears to be the optimum choice for the buffer region, since this region is relatively nonpolarizable and far from the protonation site.

The last entry in Table 2 indicates a structural collapse of the *ab initio* region onto the EFP region. This is presumably due to the lack of a repulsive potential combined with a rather unphysical division of the molecule into EFP/buffer/*ab initio* regions, due to the small size of the molecule. Next, we consider a larger system.

Further tests of the EFP/buffer/*ab initio* method were performed by computing the $\epsilon-$N PAs of two different conformations of the Gly-Lys-Gly tripeptide: one with an intramolecular hydrogen bond and one without. The latter undergoes a larger conformational change in the EFP region and is therefore a more stringent test. The results are summarized in Tables 3 and 4, respectively.

The first columns of both tables show that the presence of the intermolecular hydrogen bond reduces the effect of conformational rearrangement on the PA by 60−70%. This is a promising result, given the large number of intermolecular hydrogen bonds in proteins.

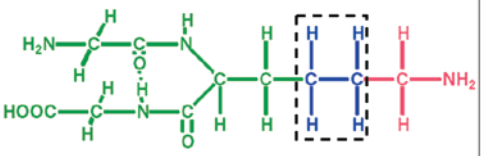
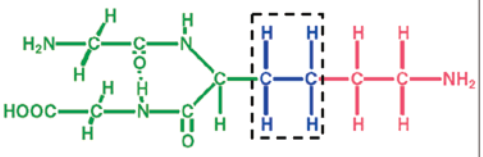
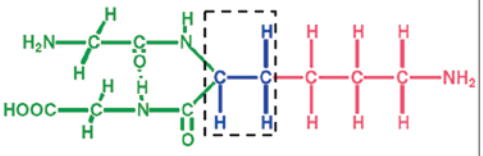
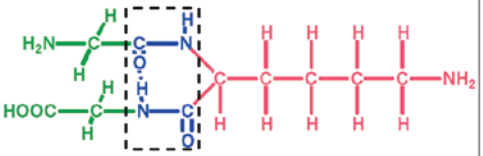
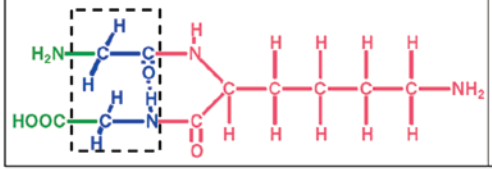
The second columns of both tables demonstrate that molecular environment can have a significant effect (up to 7.2 kcal/mol) on the PA of lysine. The effect is larger for the hydrogen-bonded conformation, despite the fact that it undergoes a smaller conformational change. The environmental effects are largely captured by the EFP representation, as shown by the data in the last columns. Again, the optimum choice of buffer region is the [$\alpha\beta$]-buffer, which for both conformations reduces the error to below 0.5 kcal/mol relative to the constrained all-*ab-initio* calculation. Moving the buffer region further out on the backbone increases the absolute error to 0.8−2.0 kcal/mol, presumably since that region is more polarizable. Thus, since the EFP region is polarizable, it is not necessarily a worse representation of the charge density than the all-*ab-initio* buffer region.

All molecules discussed here are small enough to allow full *ab initio* calculations, to gauge the accuracy of the new methodology. This also eased the construction of the EFPs, since they could be derived from a single *ab initio* calculation. We are currently applying the EFP/buffer method to larger proteins (Turkey ovomucoid third domain, α-chymotrypsin, and alcohol dehydrogenase) where the EFPs must be constructed from a series of *ab initio* calculations on smaller overlapping pieces, rather than on the entire structure. We will report on these first principles hybrid calculations on proteins in a future paper.

D. Interface with the Continuum. In the discrete approach to solvation, the size of the solvent configuration space increases dramatically with the number of solvent molecules, so this can become a computational bottleneck. In the dielectric continuum model, the solvent is described as an infinite, isotropic dielectric in which the solute is embedded. In this model, the solute polarizes the solvent via its dielectric constant. The solvent in turn polarizes the solute. The "reaction field" of the solvent is calculated either analytically or numerically depending upon the complexity of the electrostatic problem. Various cavities have been considered, ranging from regular shapes like spheres and ellipsoids to molecular shapes such as a cavity constructed from interlocking spheres surrounding the atoms of the solute.[1]

Since the continuum model neglects the specific interactions between solute and solvent molecules, it is desirable to develop a model that includes both discrete interactions between solute and nearby solvent molecules and the average interaction between solute and solvent molecules that are further away from the solute. Such a model would incorporate both discrete and continuum descriptions of the solvent. A tractable approach would be to use three different layers to describe the system, in which the solute, perhaps plus a small number of solvent molecules, is described using *ab initio* quantum mechanics, with the remaining explicit solvent molecules treated with a model potential. The outer layer would be a continuum. Van Duijnen

TABLE 3: Same as in Table 2, for H-Bonded Gly-Lys-Gly Tripeptide in kcal/mol[a]

	1 Reference ab initio calculation	2 QM/buffer calculation	3 QM/buffer/EFP calculation
	231.9 0.0	239.1 +7.2	234.1 +2.2
	231.8 0.0	237.3 +5.5	232.5 +0.7
	231.8 0.0	237.8 +6.0	232.2 +0.4
	231.6 0.0	234.0 +2.4	232.9 +1.3
	231.4 0.0	233.5 +2.1	230.4 −1.0

[a] The proton affinity of the fully relaxed tripeptide is 231.0 kcal/mol.

and co-workers explored this idea and developed a three-layer model[3c] by representing the explicit solvent molecules as a combination of point charges and atomic polarizabilities. Here, we summarize a three-layer model in which the second layer of solvent molecules is represented by effective fragment potentials.

The first continuum method that has been interfaced with the EFPs is the simple self-consistent reaction field (SCRF) with a spherical cavity (Onsager method). Only terms up to dipole are taken into account in the interaction energy between solute and solvent. The combined model is referred to as EFP + Onsager. Interfacing the EFP method with the polarizable continuum model (PCM)[1a] is in progress. The EFP + Onsager method is summarized below.

The electrostatic interaction energy between a solute inside a sphere and the surrounding solvent can be written[1] as

$$E_{\text{int}} = -\frac{1}{2}\sum_{l}\sum_{m=-l}^{l}\frac{(l+1)(\epsilon-1)}{[l+\epsilon(l+1)]a^{2l+1}}M_l^m M_l^m \qquad (41)$$

where a is the radius of the spherical cavity, ϵ is the dielectric constant and M_l^m is the expectation value of the multipole moment m_l^m of the solute:

$$M_l^m = \int \mathrm{d}\tau\, \rho m_l^m$$

$$m_l^m = \left(\frac{4\pi}{2l+1}\right)^{1/2} r^l Y_l^m(\theta,\phi)$$

where $Y_l^m(\theta,\phi)$ and ρ are the spherical harmonics and the charge density of the solute, respectively. Extracting the first two terms from eq 41 gives

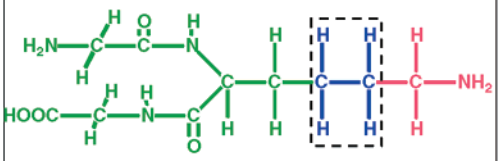
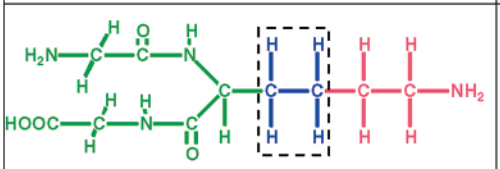
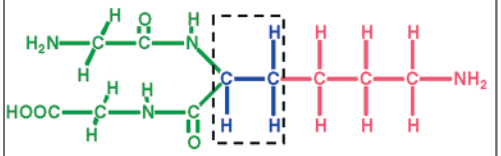
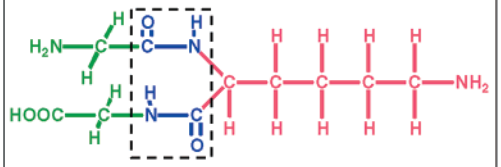
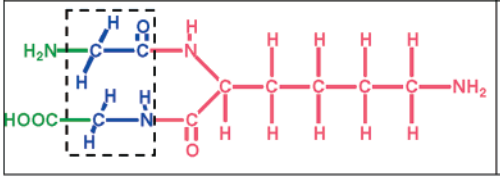
$$E_{\text{int}}^{(1)} = -\frac{1}{2}\left(1-\frac{1}{\epsilon}\right)(1/a)Q^2$$

and

$$E_{\text{int}}^{(2)} = -\frac{1}{2}\frac{2(\epsilon-1)}{(2\epsilon+1)}\frac{1}{a^3}\mu^2$$

where Q and μ are the charge and dipole moment of the solute,

TABLE 4: Same as in Tables 2 and 3, but for Non-H-Bonded Gly-Lys-Gly Tripeptide[a]

	1 Reference ab initio calculation	2 QM/buffer calculation	3 QM/buffer/EFP calculation
	237.1 0.0	239.5 +2.4	239.3 +2.2
	237.0 0.0	237.4 +0.4	237.7 +0.7
	237.0 0.0	237.7 +0.7	237.3 +0.3
	236.7 0.0	237.9 +1.2	237.5 +0.8
	236.5 0.0	236.6 +0.1	234.5 –2.0

[a] The proton affinity of the fully relaxed tripeptide is 235.2 kcal/mol.

respectively. In the Onsager model only these two terms are considered. The energy of the system is then given by

$$E = E_0 - \frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)(1/a)Q^2 - \frac{1}{2}g\mu^2 \qquad (42)$$

where $g = [2(\epsilon - 1)/(2\epsilon + 1)]\ 1/a^3$ and E_0 is the energy of the quantum mechanical solute plus the fragments. The second term on the right-hand side of eq 42 is additive. So, the energy functional in the EFP + Onsager model is given by

$$L = E_0 - \frac{1}{2}g\mu^2 - W(\langle\Psi|\Psi\rangle - 1) \qquad (43)$$

where W is a Lagrange multiplier ensuring normalization of the wave function. The total dipole moment of the system μ, has three contributions:

$$\mu = \mu_{ab} + \mu_{\text{efp}} = \mu_{ab} + \mu^s_{\text{efp}} + \mu^i_{\text{efp}} \qquad (44)$$

The three terms on the right-hand side of eq 44 represent the dipole moment of the *ab initio* part, the static dipole moment of the fragments and the induced dipole moment of the fragments, respectively. Taking the variation of the functional in eq 43 with respect to the wave function parameters and setting it to zero gives

$$\delta L = \delta E_0 - g\mu\delta\mu - W\delta\langle\Psi|\Psi\rangle + \text{cc} = 0 \qquad (45)$$

where "cc" denotes the complex conjugate of the terms given. Explicit expressions for δE_0 are given in ref. **4**. Here, we focus our attention on the second term on the right-hand side of eq 45.

$$\delta\mu = \delta\mu_{ab} + \delta\mu_{\text{efp}} = \delta(\langle\Psi|\hat{\mu}_{ab}|\Psi\rangle + \mu_{\text{nuclear}}) + \delta(\sum_j \alpha_j(\langle\Psi|\hat{F}_j|\Psi\rangle + F_{j,\text{nuclear}} + F^s_{j,\text{efp}} + F^i_{j,\text{efp}})) \qquad (46)$$

where the α_j are the polarizabilities of the fragments defined at the points j. $\hat{F}_j$ is the *ab initio* electric field operator, and, $F_{j,\text{nuclear}}$, $F^s_{j,\text{efp}}$, $F^i_{j,\text{efp}}$ are the electric fields due to the *ab initio* nuclei, static dipole of the fragments and induced dipole of the fragments, respectively. Taking the variations in eq 46 explicitly and inserting the result into eq 43 gives the following Schrodinger equation

$$[H_0 - g\mu.\hat{\mu}_{ab} - g\mu\cdot\sum_j \alpha_j\hat{F}^{\text{el}}_j]|\Psi\rangle = W|\Psi\rangle \qquad (47)$$

where μ is the total dipole moment of the system as defined in eq 44. $\hat{\mu}_{ab}$ is the dipole moment operator for the *ab initio* part.

For a neutral molecule the total energy of the system is given by

$$E = W + \frac{1}{2}g\mu\mu - g\mu\mu^{s}_{\mathrm{efp}} - g\mu[\sum_j \mu_j^{\mathrm{tot}} - \sum_j \alpha_j \langle\Psi|\hat{F}_j^{\mathrm{el}}|\Psi\rangle] \quad (48)$$

where μ_j^{tot} is the total induced dipole moment of the fragments at point j. As discussed previously,[4] the expectation value of the functional in eq 43 is not equal to the energy given in eq 42 due to the nonlinear nature of the Schrodinger equation. This is reconciled by adding the extra terms to W in eq 48 to get the total energy. The above equations have been coded and implemented in the electronic structure code GAMESS.[14]

III. Applications

A. Water Clusters. Understanding the intermolecular interactions in water is of great interest due to its importance in biological systems and as a solvent in synthesis and separation processes. Furthermore, the understanding of water clusters is a key step in linking molecular properties to bulk behavior. In separate studies water clusters for $(H_2O)_n$ of small ($n = 3-5$)[6] and moderate ($n = 6-20$)[7] size were examined using the EFP1 method. The structures, relative energetics and isomerization barrier heights for the smaller clusters compare quite favorably with the corresponding *ab initio* Hartree−Fock results that had been published previously by Wales and Walsh.[27] The study of the larger clusters serves as both a more extensive test of the method and an evaluation of the efficacy of combined Monte Carlo/simulated annealing methods for identifying global minima in such clusters. Here, the clusters of six and 20 waters are briefly summarized to illustrate the salient features of that work.

Simulated Annealing (SA) methods were tested by starting from at least four randomly chosen geometries for each cluster size. Random geometries were obtained by carrying out Monte Carlo steps at a high temperature (25000 K) for greater than 1000 steps. Since overly compressed geometries cause problems in converging the self-consistent polarizabilities in the EFP, geometries were discarded if two atoms from different molecules were separated by less than 1.3 angstroms.

The most successful method studied was the Monte Carlo with minimization method of Li and Scheraga,[28] implemented in a manner similar to that of Wales and Hodges.[29] The minima found by the SA algorithm were confirmed by carrying out Hessian calculations at the EFP level of theory and verifying that there were no imaginary vibrational frequencies. Full *ab initio* calculations, all using the polarized double-ζ basis set of Dunning and Hay[30] (DH(d,p)), were carried out on the configurations found to be minima on the EFP potential energy surface (PES). Restricted Hartree−Fock (RHF) geometry optimizations were carried out, and single point energies were calculated with Møller−Plesset second-order perturbation theory (MP2)[31,32] at the RHF geometries. For the water hexamer, geometry optimizations were also carried out at the MP2 level, and single-point coupled-cluster (CCSD[T])[33−36] energies were obtained at the MP2 geometries. All calculations were carried out with the GAMESS[14] program except for the CCSD(T) energies, which were calculated with the ACES II[37] program using the D95* [38] basis set.

1. Hexamer. The energies for six minima on the potential energy surface of $(H_2O)_6$ are given in Table 5. These structures, shown in Figure 3, have been proposed previously[39−41] as the likely candidates for the equilibrium structure.

Four of the six configurations are predicted to be the global minimum by one or more levels of theory. However, it seems likely that the global minimum is either the cage structure, as was found by Wales and Hodges[29] at the TIP4P level, or the prism structure, which is the minimum in all four previous studies that used the MP2 level of theory.[41,43−45] The prism is predicted to be the global minimum by EFP, MP2 and CCSD(T). In fact, the predicted energy ordering of the six structures is the same at these three levels of theory. However, all of these minima are close in energy, and only two structures, the bag

TABLE 5: Binding Energies (kcal/mol of $(H_2O)_6$), Contributions to the EFP, Relative Energies, and Dipole Moments (μ, Debye) for $(H_2O)_6$[a]

	prism	cage	book	bag	cyclic	boat
Binding Energies						
EFP	−42.42	−41.90	−41.45	−40.61	−41.14	−40.09
RHF	−42.86	−42.49	−42.44	−41.58	−43.10	−42.12
MP2//RHF	−55.86	−55.06	−54.13	−53.27	−53.75	−52.36
MP2//MP2	−58.25	−57.52	−56.49	−55.65	−55.75	−54.29
CCSD(T)//MP2	−55.10	−54.30	−53.10	−52.20	−52.20	−50.80
no. of H-bonds	9	8	7	7	6	6
electrostatic	−55.76	−55.33	−54.67	−53.46	−53.47	−51.77
repulsion	22.77	23.08	23.77	23.41	23.83	22.66
polarization	−9.44	−9.65	−10.55	−10.57	−11.51	−10.99
Energies per Hydrogen Bond						
energy (EFP)	−4.71	−5.24	−5.92	−5.80	−6.86	−6.68
electrostatic	−6.20	−6.92	−7.81	−7.64	−8.91	−8.63
repulsion	2.53	2.88	3.40	3.34	3.97	3.78
polarization	−1.05	−1.21	−1.51	−1.51	−1.92	−1.83
Relative Energies						
EFP	0.00	0.52	0.98	1.81	1.28	2.33
RHF	0.00	0.37	0.42	1.29	−0.24	0.75
MP2//RHF	0.00	0.81	1.74	2.60	2.11	3.50
MP2//MP2	0.00	0.73	1.75	2.60	2.49	3.95
CCSD(T)//MP2	0.00	0.80	2.00	2.90	2.90	4.30

[a] EFP, Effective fragment potential; RHF, restricted-spin Hartree−Fock with DH(d,p)[20] basis set; MP2//RHF, MP2[21,22] single-point energies using DH(d,p) basis at RHF/DH(d,p) optimized geometries; MP2//MP2, optimized at MP2 level using DH(d,p) basis; CCSD(T)//MP2, coupled cluster energies at MP2 optimized geometry.

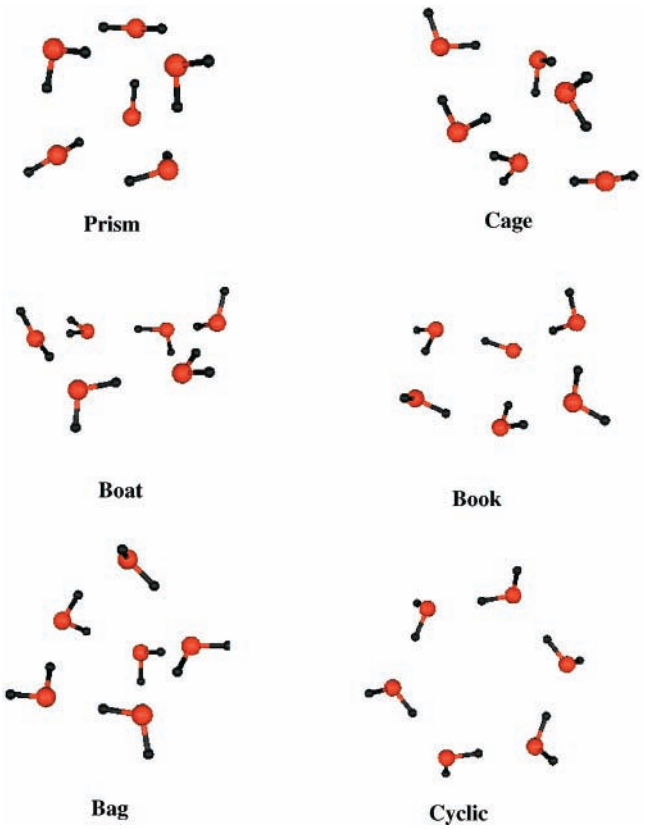


Figure 3. Minimum energy structures for $(H_2O)_6$.

TABLE 6: Binding Energies for $(H_2O)_{20}$ (kcal/mol of Clusters)

$(H_2O)_{20}$	EFP	RHF	MP2
p3a	−191.19	−189.92	−248.39
$(D_{2d})_4$	−188.83	−187.33	−247.04
pps-b	−190.49	−188.72	−247.03
pps-c	−190.42	−188.63	−246.90
$(S_4)(D_{2d})_3$	−188.63	−187.11	−246.88
pps-d	−188.72	−187.90	−246.44
$(D_{2d}S_4)_2$	−188.35	−186.77	−246.39
p3b	−189.17	−187.97	−246.37
pps-e	−188.99	−187.72	−246.14
pps-f	−189.04	−187.45	−245.68
pps-a	−188.90	−187.31	−245.63
$(S_4)_4$	−188.12	−186.24	−245.57

and the boat, are clearly not the global minimum. At the Hartree−Fock level of theory, the cyclic structure is the global minimum.

The agreement between the EFP and RHF binding energies is quite good (≤1.0 kcal/mol) for the prism, cage, book, and bag structures. For the cyclic and boat structures, the difference is 2.0 kcal/mol. The MP2 and CCSD(T) results indicate that the prism is the lowest energy structure on the potential energy surface. The CCSD(T) binding energies are consistently about 3 kcal/mol smaller than those predicted by MP2, but the relative energies predicted by the two methods are quite similar. The much larger binding energies predicted by the correlated methods reflect the omission of dispersion contributions in both the RHF and EFP1 methods.

The EFP electrostatics, repulsion, and polarization energies (total and per H-bond) are given in Table 5. While the 3-dimensional prism and cage structures have more H-bonds, these bonds are weakened relative to those in the more 2-dimensional cyclic and boat structures due to a distortion from the preferred linear orientation. The 2-D structures have fewer H-bonds, but these bonds are less distorted and thus shorter and stronger. The book and bag structures have intermediate numbers of H-bonds and moderate degrees of H-bond distortion. Competition between these two factors results in the six structures having similar energies. This tradeoff is unique to the water hexamer. For smaller clusters, the H-bond distortion is too large in 3-D structures, and thus the cyclic structures are lowest in energy. For larger clusters, the global minima are clearly 3-dimensional structures. The stronger H-bonds in the cyclic and boat structures have larger average components of electrostatic, repulsion, and polarization than their weaker counterparts in the prism and cage structures, but the total electrostatic interaction is greater in the 3-D prism and cage structures due to their greater numbers of H-bonds. The molecules in the cyclic, boat, and book structures are so much more polarized than those in the 3-D prism and cage structures that the total polarization energies in these three clusters are greater despite having fewer H-bonds. As a result, the combined electrostatic plus polarization interactions are nearly equivalent for the prism, cage, book, and cyclic structures.

2. n = 20. For the $(H_2O)_{20}$ system, basin hopping was carried out at thirty-five temperatures with 2400 geometries per temperature for a total of 84000 trial geometries. For each of these three systems, four simulations were carried out at each of the four starting geometries for a total of sixteen production runs.

Table 6 lists the interaction energies for the minima found for $(H_2O)_{20}$ (see ref 7 for the actual structures). Tsai and Jordan[42] identified three nearly degenerate fused cube structures for the lowest energy configurations of $(H_2O)_{20}$, while Wales and Hodges[29] found another structure (p3a) about 1 kcal/mol lower in energy on the TIP4P PES. The latter structure is more than 2 kcal/mol lower in energy on the EFP and RHF surfaces, and 1.3 kcal/mol more stable in the MP2 calculations. Two other low-energy minima, pps-b and pps-c, are just 0.7 and 0.8 kcal/mol, respectively, above the EFP global minimum. The MP2 structures are nearly isoenergetic with two fused cube structures: the $(D_{2d})_4$ structure identified as the global minimum by Tsai and Jordon;[42] and $(S_4)(D_{2d})_3$ not identified in the previous studies. Once again, the EFP method reproduces the RHF energetics to within about 1.5 kcal/mol. In general, the relative energies predicted by the effective fragment potential, RHF and MP2 methods are all in quite good agreement. Of course, the uncorrelated methods underestimate the binding energies.

The individual components of the interaction energy obtained with the EFP have facilitated the analysis of the balance between the number of hydrogen bonds and the strain in the hydrogen bonds. The anisotropic nature of the electrostatic interaction leads to preferred bond angles in H-bonding, and thus makes electrostatics the most important factor in determining the stability. Increases in the magnitudes of the polarization and repulsion terms are often the result of the shorter bond lengths in structures with more favorable electrostatics. However, the repulsive term is on average larger whenever the molecules are closer, and thus can be more important in single prism configurations than in multiple fused-prism geometries. This renders single prism configurations less stable. Because of the large role that electrostatics plays in water−water interactions, the EFP can be a powerful tool in the prediction and analysis of water cluster structures.

B. Menshutkin Reaction. The Menshutkin reaction[46] is an important test of any solvation method, since it involves the reaction of two neutral species, an alkyl halide and an alkylamine, to produce an associated ion pair (ammonium salt) and eventually separated ions:

$$RX + R'NH_2 \rightarrow [X]^-[R'(R)NH_2]^+ \rightarrow X^- + [R'(R)NH_2]^+$$

In the gas phase, of course, this process is rather endothermic, but the product ion pair and separated ions are both expected to be stabilized by polar solvents. There have been several theoretical studies of the Menshutkin reaction, using various theoretical methods, all with R = CH_3 and R′ = H. Sola and co-workers[47] used the modest 3-21G basis set and X = Br to describe the solute plus two waters, then surrounded this cluster with a continuum to predict a free energy of activation of 8.3 kcal/mol. Gao and co-workers[48] used a semi-empircial QM/TIP3P MM approach with X = Cl, and predicted a 26.3 kcal/mol free energy of activation. Rivail[49] used a continuum method for X = Cl to obtain a free energy of activation that is similar to that predicted by Gao et al. The only quantitative experimental data comes from the Okamoto group[50] who find a free energy of activation of 23.5 kcal/mol for R = CH_3, R′ = H, X = I.

The EFP1 method has been used to study the Menshutkin reaction for R = CH_3, R′ = H, X = Br.[51] The solute system was treated with both restricted Hartree−Fock (RHF) and second-order perturbation theory, using a polarized double-ζ basis set with diffuse functions added to Br. The effects of electron correlation are found to be very small for the gas-phase reaction,[51] so only RHF wave functions were used for the solvated systems. For two waters, full geometry optimizations were performed using both *ab initio* and EFP waters, including identification of the transition state and determination of the minimum energy path (MEP) that connects reactants (reactant molecule-pair) with product ion pair. For larger numbers of

TABLE 7: Calculated Energies (kcal/mol) at 0 K, Relative to the Lowest Energy Molecule-Pair Reactant in the Menshutkin Reaction as a Function of the Number of Solvating Water Molecules[a]

no. of waters			ΔE (kcal/mol) EFP	RHF
0		REAC		0.0
		TS		34.0
		PROD		−4.7
2		REAC	0.0	0.0
		TS	22.5	22.2
		PROD	−19.4	−19.2
4	A	REAC	3.9	3.5
		TS	17.6	16.9
		PROD	−30.7	−31.2
4	B	REAC	0.0	0.0
		TS	18.0	17.5
		PROD	−26.2	−26.3
6	A	REAC	0.9	0.0
		TS	23.0	20.3
		PROD	−19.6	−23.5
6	B	REAC	0.0	0.0
		TS	16.0	15.6
		PROD	−29.1	−28.8
8		REAC	0.0	0.0
		TS	20.3	19.4
		PROD	−31.4	−31.3

[a] Zero-point energies and temperature effects are not included.

waters, geometries were optimized using EFP waters, and RHF single point energy calculations were then carried out at the EFP geometries. Since the primary goal of this study was to assess the ability of the EFP method to reproduce the key features of the Menshutkin reaction, the configuration space was not exhaustively searched for a given number of water molecules. Rather, several plausible arrangements of waters were chosen as starting points for geometry optimizations and saddle point searches.

It is important to emphasize here that *no new EFP parameters were employed for this analysis.* That is, the EFP1 parameters that were determined as described in section IIA are subsequently used for all applications, with no further adjustment.

The energetics of the Menshutkin reaction as a function of the number of solvating water molecules are summarized in Table 7. Note that REAC and PROD in this table refer to the molecule-pair complex in the entrance channel and the ion-pair product in the exit channel, so there are no basis set superposition error issues. Clearly, the complete separation into ions is highly endothermic in the gas phase.[51] The entries "A" and "B" in the table for four and six waters refer to two different reaction paths for these numbers of waters.

Several points regarding this table are noteworthy. First, as the number of water molecules increases, the net exothermicity increases and the barrier height decreases, as one would expect for a polar solvent. The differences between the two different reaction paths for four and six waters is small, but not insignificantly so. One would therefore expect that a careful statistical sampling of the configuration space will become increasingly important as the number of water molecules increases. This is very likely the reason that the barrier appears to increase slightly upon increasing the number of waters from six to eight. The agreement between the energetics predicted by the EFP method and those from the all *ab initio* calculations are excellent. The two sets of relative energies and barrier heights generally agree to within 1 kcal/mol, although the difference is somewhat larger for the six water (A) case. This suggests that the EFP method will be similarly successful in a variety of solvation problems.

TABLE 8: ΔG (298.15 K) of Activation and Reaction (kcal/mol), Calculated at the RHF/DZVP Level for the Menshutkin Reaction with 0, 2, and 8 Solvating Water Molecules

no. of waters		ΔG_{298} (kcal/mol) EFP	AI
0	REAC		0.0
	TS		40.6
	PROD		2.8
2	REAC	0.0	0.0
	TS	28.3	27.7
	PROD	−10.2	−10.0
8	REAC	0.0	
	TS	22.7	
	PROD	−23.8	

TABLE 9: Comparison of Wall Clock Times (s, on IBM RS6000/350) for RHF vs EFP[a]

no. of waters	wall clock RHF	EFP	Δ (wall clock) RHF	EFP
2	4006 (5399)	1376 (2118)	3117 (3672)	487 (391)
4	11768 (17905)	1658 (2422)	7762 (12506)	282 (304)
6	(26684)	2054 (2981)	(8779)	396 (559)
8	(45029)	1752 (3134)	(18345)	221 (153)

[a] Times in parentheses are from direct SCF;others are for conventional SCF.

The corresponding free energies, calculated by constructing the harmonic oscillator/rigid rotor partition functions, are listed in Table 8 for zero, two and eight waters. The overall trends in this table are similar to those discussed above, including the excellent agreement between EFP and RHF methods. For the largest number of waters, the predicted free energy of activation is 22.7 kcal/mol. This is in good agreement with both experiment and the earlier theoretical results of the Gao and Rivail groups.

Finally, a sense for the relative computer times required for the EFP vs RHF calculations may be found in Table 9. These calculations were performed on an RS6000/350 workstation, and only wall clock times are presented. The most telling comparison is that for the increment in the time required for the calculation as the number of water molecules is increased. For the EFP calculations, this increment is a direct assessment of the cost of the EFP computations, since the same RHF solute is present for each calculation. The EFP increment is essentially constant as two waters are successively added. The observed variation is due to small variations in the number of SCF iterations that are required for convergence. The incremental cost of the RHF calculations is more than 2 orders of magnitude larger and is increasing as n increases.

C. Glycine Neutral/Zwitterion Equilibrium. As an example of the interplay between discrete and continuum treatments of solvation, we consider the equilibrium between neutral (N) and zwitterionic (Z) glycine, the simplest amino acid. In the gas phase, one expects the neutral isomer to be more stable, while the zwitterion is the global minimum in aqueous solution.[52] In the present study, the geometries of bare N and Z glycine were optimized in solution with the 6-31++G** basis set[53] using the Onsager self-consistent reaction field (SCRF) model. The cavity radii for the optimizations were taken as 3.62 and 3.74 Å for N and Z, respectively. These radii were obtained by

TABLE 10: $E(N) - E(Z)$ (kcal/mol)

gas phase (RHF/6-31++G**)	−30		
discrete		continuum only	
EFP (RHF/6-31++G**)	−4.6	SCRF/6-31++G**	−9.0
ab initio (RHF/6-31G*)	−3.2	SM5.42R/6-31G*	−1.0
		IPCM/6-31++G**	−3.2

Discrete + Continuum

method	energy difference
EFP + SCRF (6-31++G**)	−5.8
AI + SCRF(6-31G*)	−4.3
SM5.42R (6-31G*)	−1.1
IPCM (6-31G*)	8.6

calculating the molecular volume using GAUSSIAN 94.[54] Two other solvation models, IPCM[1b]/6-31++G** and SM5.42R[1c]/6-31G*, were also used to calculate the energies of these two optimized structures. Second, $N(H_2O)_8$ and $Z(H_2O)_8$ were optimized using effective fragment waters with the RHF/6-31++G** method for N and Z. Finally, the effect of the bulk solvent was taken into account by embedding the clusters in a continuum. The Onsager continuum was used with the EFP waters, while *ab initio* waters at the EFP geometries were used when the continuum was described by the IPCM and SM5.42R models. The calculations were performed using GAMESS,[14] GAUSSIAN94[54] and GAMESOL.[55]

Two issues that arise in a supermolecule-continuum calculation are the number of solvent molecules to be used in the calculation, and how best to sample the nuclear degrees of freedom of the explicit solvent molecules. With regard to the first question, adding waters to the hydrophilic parts of the solute should give a reasonable approximation to the first solvation shell. It is found that at least eight water molecules are necessary to fill hydrophilic areas of glycine and act as a 'first solvation shell'. In the present work, eight explicit water molecules are included. The configuration space of the eight water molecules was not exhaustively sampled. Instead, full geometry optimizations were performed for reasonably chosen glycine−water cluster structures.

The energy differences between the optimized N and Z isomers using the Onsager, SM5.42R and IPCM methods are shown in Table 10, as well as the gas phase energy difference. The gas phase energy difference was calculated using the Onsager optimized geometry of Z. A smaller basis set was used for the SM5.42R model, since it has not been parametrized for the larger basis. All three continuum methods cause a significant stabilization of Z relative to N, but all three still find N to be slightly more stable than Z. One reason for these results is that the three models assume that the polarization of the solvent is linearly proportional to the electric field of the solute. However, for the zwitterion, nonlinear effects in the polarization of the solvent become important due to the charges present at the two ends of the solute.[56,57]

Next, $N(H_2O)_8$ and $Z(H_2O)_8$ were optimized using RHF/6-31++G** for glycine with EFP waters. The hydrophilic sites of both N and Z are filled by the explicit waters. Analogous full RHF/6-31G* calculations were also carried out. Table 10 shows the results of these calculations. Note that EFP and *ab initio* results differ by only 1.4 kcal/mol. The cluster calculations stabilize Z relative to N, but $N(H_2O)_8$ is still predicted to be more stable than $Z(H_2O)_8$. In the gas phase, the RHF/6-31++G** dipole moment of Z (13.4 D) is much higher than that of N (1.3 D). But, $Z(H_2O)_8$ has a smaller dipole moment (3.5D) than $N(H_2O)_8$ (5.3 D). Since the Onsager model considers only the dipole moment of the solute, $N(H_2O)_8$ will clearly be more stabilized than $Z(H_2O)_8$ in this model, contrary to the stabilization of bare N and Z in the Onsager model.

Next, the combined cluster/continuum effect of the bulk solvent is taken into account by surrounding the clusters with a continuum, using the EFP + Onsager model, as well as all three continuum models with *ab initio* waters. The results are shown in Table 10. As suggested above, the Onsager model stabilizes the N cluster more than the Z cluster due to the larger $N(H_2O)_8$ dipole moment. $N(H_2O)_8$ is more stable than $Z(H_2O)_8$ by 5.8 kcal/mol (without the continuum the energy difference is 4.6 kcal/mol). The relative stabilization of N using 8 *ab initio* waters is similar. In the SM5.42R model, N and Z clusters are very close in energy, though N is still more stable by 1.1 kcal/mol. The SM5.42 continuum stabilizes the $Z(H_2O)_8$ cluster by 2.1 kcal/mol. The IPCM model favors the Z cluster by 8.6 kcal/mol. So, the IPCM continuum stabilizes the $Z(H_2O)_8$ cluster by ∼11.8 kcal/mol. A more definitive study of these effects would involve a systematic sampling of configuration space as a function of the number of water molecules.

An approximate sense for the level of agreement with experiment can be found by comparing the SM5.42 and IPCM values with the experimental free energy change, N → Z in solution, of 7.67 kcal/mol favoring Z.[20] Although the IPCM model takes into account only the electrostatic part of the solvation free energy, these results clearly suggest that the combination of a small number of explicit waters and a sophisticated continuum model will be an effective tool for the study of N−Z equilibria.

IV. Summary and Future Directions

The effective fragment potential (EFP1) method has now been applied successfully to a wide array of problems, including the analysis of water cluster structures and the effects of aqueous solvation on chemical reactions in ground and electronically excited states. The principle limitation of the method, at present, is that correlation contributions, most notably dispersion, are not yet included in the method. A careful study of the water dimer potential energy surface using both second-order perturbation theory and coupled cluster CCSD(T) calculations has been completed,[58] and these calculations will form the basis for the development of the dispersion term.

A bottleneck in the EFP1 method is that the extension of the method to solvents other than water requires the determination of dimer potential energy surfaces for each solvent of interest, followed by a fitting procedure for the exchange repulsion/charge-transfer term. Therefore, the more appealing approach for generalization of the method is to avoid adjustable fitted parameters by deriving these latter terms, as well as the dispersion, from first principles. This is the guiding philosophy of the EFP2 method. Generalized formulations for both the exchange repulsion and charge penetration, as well as EFP components of the energy gradient, have already been derived and implemented. Analogous derivations for dispersion and charge transfer and the *ab initio*/EFP contribution to the gradient are in progress.

A particularly exciting development is the derivation of the EFP/*ab initio* link across covalent bonds. While this approach was initially developed to study the biochemistry of peptides and enzymes, the potential applications are clearly much broader. Since the method incorporates most of the important physical interactions directly from quantum mechanics and requires no adjustable parameters, it will be an obvious alternative to the MM in any QM/MM calculations.

Acknowledgment. The work described in this paper was supported by grants to M.S.G. from the National Science Foundation and the Air Force Office of Scientific Research. J.H.J. acknowledges support from the University of Iowa Biosciences Initiative Pilot Program, the Research Corporation and the Petroleum Research Fund. The contributions of Drs. Wei Chen, Paul Day, Morris Krauss, Grant Merrill, Michael Schmidt and Simon Webb are gratefully acknowledged.

References and Notes

(1) See, for example: (a) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486; (b) Tomasi, J.; Perisco, M. *Chem. Rev.* **1994**, *94*, 2027. (c) Cramer, C. J.; Truhlar, D. G. *Reviews in Computational Chemistry*; Boyd, D. B., Lipkowitz, K., Eds.; VCH: New York, 1995; Vol. 6. (d) Cramer, C. J.; Truhlar, D. G. *Solvent effects and Chemical Reactivity*; Tapia, O., Bertran, J., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1996.

(2) Levy, R. M.; Kitchen, D. B.; Blair, J. T.; Krogh-Jespersen, K. *J. Phys. Chem*. **1990**, *94*, 4470.

(3) See, for example: (a) Ten-no, S.; Hirata, F.; Kato, S. *Chem. Phys. Lett*. **1993**, *214*, 391. (b) Russell, S. T.; Warshel, A. *J. Mol. Biol*. **1985**, *185*, 389. (c) De Vries, A. H.; Van Duijnen, P. Th.; Juffer, A. H.; Rullmann, J. A. C.; Dijkman, J. P.; Merenga, H.; Thole, B. T. *J. Comput. Chem*. **1995**, *16*, 37.

(4) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys*. **1996**, *105*, 1968.

(5) Chen, W.; Gordon, M. S. *J. Chem. Phys*. **1996**, *105*, 11081.

(6) Merrill, G. N.; Gordon, M. S. *J. Phys. Chem. A* **1998**, *102*, 2650.

(7) Day, P. N.; Pachter, R.; Gordon, M. S.; Merrill, G. N. A Study of Water Clusters Using the Effective Fragment Potential and Monte Carlo Simulated Annealing. *J. Chem. Phys*. **2000**, *112*, 2063.

(8) (a) Krauss, M. *Comput. Chem.* **1995**, *19*, 199. (b) Krauss, M.; Wladowski, B. D. *Int. J. Quantum Chem.* **1998**, *69*, 11. (c) Krauss, M.; Webb, S. P. *J. Chem. Phys*. **1997**, *107*, 5771.

(9) Minikis, R. M.; Kairys, V.; Jensen, J. H. *J. Phys. Chem. A*, submitted.

(10) Bandyopadhyay, P.; Gordon, M. S. *J. Chem. Phys*. **2000**, *113*, 1104.

(11) Kairys, V.; Jensen, J. H. *J. Chem. Phys. A* **2000**, *104*, 6656.

(12) Wladkowski, B. D.; Krauss, M.; Stevens, W. J. *J. Am. Chem. Soc*. **1995**, *117*, 10537.

(13) (a) Stone, A. J. *Chem. Phys. Lett*. **1981**, *83*, 233. (b) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, U.K., 1996.

(14) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., Jr. *J. Comput. Chem.* **1993**, *14*, 1347.

(15) Jensen, J. H. *J. Chem. Phys*. **1996**, *104*, 7795.

(16) Jensen, J. H.; Gordon, M. S. *Mol. Phys*. **1996**, *89*, 1313.

(17) Jensen, J. H.; Gordon, M. S. *J. Chem. Phys*. **1998**, *108*, 4772.

(18) Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. *J. Chem. Phys*. **2000**, *113*, 7300.

(19) Kairys, V.; Jensen, J. H. *Chem. Phys. Lett*. **1999**, *315*, 140.

(20) Maseras, F.; Morokuma, K. *J. Comput. Chem*. **1995**, *16*, 1170.

(21) Shoemaker, J.; Burggraf, L. W.; Gordon, M. S. *J. Phys. Chem*. A **1999**, *103*, 3245.

(22) Assfeld, X.; Rivail, J.-L. *Chem. Phys. Lett*. **1996**, *263*, 100.

(23) King, H. F.; Stanton, R. E.; Wyatt, R. E.; Parr, R. G. *J. Chem. Phys.* **1967**, *47*, 1936.

(24) Bagus, P. S.; Hermann, K.; Bauschlicher, C. W., Jr. *J. Chem. Phys.* **1984**, *80*, 4378.

(25) Stevens, W. J.; Fink, W. *Chem. Phys. Lett*. **1987**, *139*, 15.

(26) Edmiston, C.; Ruedenberg, K. *Rev. Mod. Phys.* **1963**, *35*, 457.

(27) (a) Wales, D. J. *J. Am. Chem. Soc.* **1993**, *115*, 11180. (b) Wales, D. J.; Walsh, T. R. *J. Chem. Phys.* **1996**, *105*, 6957. (c) Wales, D. J.; Walsh, T. R. *JCS Far. Trans.* **1995**, *92*, 2505. (d) Wales, D. J.; Walsh, T. R. *J. Chem. Phys.* **1997**, *106*, 7193.

(28) Li, Z.; Scheraga, H. A. *Proc.f Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611.

(29) Wales, D. J.; Hodges, M. P. *Chem. Phys. Lett*. **1998**, *286*, 65−72.

(30) Dunning, T. H. J.; Hay, P. J. *Methods of Electronic Structure Theory*; Plenum Press: New York, 1977.

(31) *Relativistic and Electron Correlation Effects in Molecules*; Dupuis, M., Chin, S., Marquez, A., Eds.; Plenum Press: New York, 1994.

(32) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett*. **1990**, *166*, 275−280.

(33) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.

(34) Bartlett, R. J.; Watts, J. D.; Kucharski, S. A.; Noga, J. *Chem. Phys. Lett.* **1990**, *165*, 513.

(35) Gauss, J.; Lauderdale, W. J.; Stanton, J. F.; Watts, J. D.; Bartlett, R. J. *Chem. Phys. Lett.* **1991**, *182*, 207.

(36) Watts, J. D.; Gauss, J.; Bartlett, R. J. *J. Chem. Phys.* **1993**, *98*, 8718.

(37) Stanton, J. F.; Gauss, J.; Watts, J. D.; Nooijen, M.; Oliphant, N.; Perera, S. A.; Szalay, P. G.; Lauderdale, W. J.; Gwaltney, S. R.; Beck, S.; Balkova, A.; Bernholdt, D. E.; Baeck, K.-K.; Rozyczko, P.; Sekino, H.; Huber, C.; Bartlett, R. J.; Almlof, J.; Taylor, P. R.; Helgaker, T.; Jensen, H. J. A.; Jorgensen, P.; Olsen, J. Quantum Theory Project, University of Florida.

(38) Dunning, T. H. *J. Chem. Phys.* **1970**, *53*, 2823.

(39) Masella, M.; Flament, J. P. *J. Chem. Phys.* **1997**, *107*, 9105.

(40) Liu, K.; Brown, M. G.; Carter, C.; Saykally, R. J.; Gregory, J. K.; Clary, D. C. *Nature* **1996**, *381*, 501.

(41) Kim, J.; Kim, K. S. *J. Chem. Phys.* **1998**, *109*, 5886.

(42) Tsai, C. J.; Jordan, K. D. *J. Phys. Chem.* **1993**, *97*, 5208.

(43) Tsai, C. J.; Jordon, K. D. *Chem. Phys. Lett.* **1993**, *213*, 181.

(44) Mhin, B. J.; Kim, J.; Lee, S.; Lee, J. Y.; Kim, K. S. *J. Chem. Phys.* **1994**, *100*, 4484.

(45) Pedulla, J. M.; Kim, K.; Jordan, K. D. *Chem. Phys. Lett*. **1998**, *291*, 78.

(46) (a) Menshutkin, N. Z. *Phys. Chem.* **1890**, *5*, 589. (b) Menshutkin, N. Z. *Phys. Chem.* **1890**, *6*, 41.

(47) Solà, M.; Lledós, A.; Duran, M.; Bertrán, J.; Abboud, J. M. *J. Am. Chem. Soc.* **1991**, *113*, 2873.

(48) Gao, J.; Xia, X. *J. Am. Chem. Soc*. **1993**, *115*, 9667.

(49) Dillet, V.; Rinaldi, D.; Bertran, J.; Rivail, J.-L. *J. Chem. Phys*. **1996**, *104*, 9437.

(50) Okamoto, K.; Fukui, S.; Shingu, H. *Bull. Chem. Soc. Jpn.* **1967**, *40*, 1920.

(51) Webb, S. P.; Gordon, M. S. *J. Phys. Chem.* **1999**, *A103*, 1265.

(52) Tortonda, F. R.; Pascual-Ahuir, J. L.; Silla, E.; Tunon, I. *Chem. Phys. Lett*. **1996**, *260*, 21.

(53) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys*. **1972**, *56*, 2257.

(54) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Peterson, G. A.; Montogomery, J. A.; Raghavachari, K.; Al-laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkra, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D.; Binkley, J. S.; Defress, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian94*; Gaussian Inc.: Pittsburgh, PA, 1995.

(55) Li, J.; Zhu, T.; Hawkins, G. D.; Chung, Y. Y.; Fast, P. L.; Liotard, D. A.; Rinaldi, D.; Cramer, C. J.; Truhlar, D. G. GAMESOL-version 2.2.4; University of Minnesota: Minneapolis, 1999; based on GAMESS.

(56) For a discussion on the nonlinear effects in the polarization of the dielectric see: Bottcher, C. J. F. *Theory of electric polarization*; Elsevier: Amsterdam, 1973.

(57) Gaffney, R. C.; Pierce, J. S. *J. Am. Chem. Soc*. **1977**, *99*, 4293.

(58) Stevens, W. J.; Gordon, M. S. Unpublished results.